

# DSSTox Log File:

## EPA Fathead Minnow Acute Toxicity Database (EPAFHM)

(last updated 15 February 2008)

**Description:** Information in this file documents the creation, review, and update process for the DSSTox EPAFHM SDF file and provides summary information on database content. The first section summarizes the process used for creating the initial DSSTox SDF files, and the quality assurance checks and procedures employed. A table providing field and data counts offers summary overview of EPAFHM file contents and chemical composition. The Log Table documents modifications and revisions to the database content or format in version updates. To obtain the most current version of this Log File and a record of any new modifications, a user should consult the DSSTox EPAFHM database page:

[http://www.epa.gov/ncct/dsstox/sdf\\_epafhm.html](http://www.epa.gov/ncct/dsstox/sdf_epafhm.html)

### QA and Development Notes for v1a:

EPAFHM SDF files underwent an extensive series of quality review checks prior to publication of initial launch versions. The original EPA MED Fathead Minnow database was obtained from the Source in dbf format, imported into Excel, and was used for cross-referencing and checking of all modified and added fields. We thank Chris Russom for invaluable assistance in various stages of development and quality review, clarifying numerous issues pertaining to the experimental data and their representation in the original database, and approving all changes to field names and contents. Chemical structures were initially obtained by conversion of SMILES codes provided in the original database using CambridgeSoft ChemOffice 2002 ChemDraw (ver 7.0 for Windows) for MS Excel. The ChemFinder website (<http://chemfinder.cambridgesoft.com/>) was used extensively for checking CAS-to-structures and for retrieving CAS numbers for parent forms of salts and complexes. CambridgeSoft ChemFinder (ver 7.0 for Windows) was also used for automatic generation of SMILES codes from structures and both ChemFinder and ACD ChemFolder (ver 6.0 for Windows) were employed for "Structure-to-Name" or "Name-to-Structure" features. Original SMILES codes provided by the Source were converted to ChemFinder-generated SMILES and verified within the ChemFinder application.

### Notes for v2a:

For version 2a, a variety of fields have been added. IUPAC systematic chemical names, **ChemName\_IUPAC**, were computed by Marc Nicklaus (NCI) using the ACD Labs IUPAC Name-Generation software (ACD/NameBatch, version 8.05). **INChI** codes were automatically generated from the final DSSTox SDF using a pre-release version of the publicly available program, wINChI11b.exe, accessible from the NIST INChI developers (<http://chemdata.nist.gov/IChI/INChIv11b.zip>). We thank Chris Hardy (Lhasa Limited, UK) for alerting us to the 14 invalid CAS provided in the original EPAFHM Source database and DSSTox EPAFHM\_v1a files, and providing us with the correct CAS for these structures. These CAS were originally unavailable to the Main Citation Source authors, so invalid CAS placeholders were used.

### Notes for v3a,b:

Revised DSSTox Standard Chemical Fields are included (see <http://www.epa.gov/ncct/dsstox/MoreonStandardChemFields.html>) along with updated InChI codes (version 1.0), recomputed IUPAC chemical names (ACDLabs ACD/Name, version 9.0), and many regenerated 2D structures with stereochemistry of steroidal compounds rendered in more standardized form. Many EPAFHM Source Toxicity Fields were also provided more descriptive names and Chemical Names were converted to lower case from all CAPS. Additionally, an extensive quality review of all DSSTox chemical records was performed, resulting in numerous corrections and modifications to chemical structures and added information (CASRN, representative structures for mixtures, etc) throughout DSSTox data files. For more information on current review procedures, see <http://www.epa.gov/ncct/dsstox/ChemicalInfQAProcedures.html>

### Notes for v4a:

EPAFHM\_v4 has no new chemical records but has several minor QA corrections, field entry revisions, field changes, etc. Changes to DSSTox Standard Chemical Fields include new ID fields: **DSSTox\_RID**, **DSSTox\_Generic\_SID** and **DSSTox\_FileID** (replacing **DSSTox\_SID** and **DSSTox\_ID\_FileName**) (see <http://www.epa.gov/ncct/dsstox/MoreonStandardChemFields.html>). Also, entries in **TestSubstance\_Description** field have been simplified. **LC50\_Ratio** field has been converted to a pure numeric field, with field notes moved to **LC50\_Note** field. Expanded text entries are provided for several Source-Specific fields (**MLOGP**, **LC50\_Note**, **MOA**, **MOA\_MixtureTest**, **FishAcuteToxSyndrome**, **FishBehaviorTest**).

#### Notes for v4b:

EPAFHM\_v4b includes minor structure changes/modifications and two new summary activity fields for use in PubChem and structure-activity relationship studies: **ActivityOutcome\_EPAFHM** (entries of active, inactive, or inconclusive) and **ActivityScore\_EPAFHM** (INTEGER[0-100]). In addition, the **MOA\_Confidence** field has expanded text entries, and the new **STRUCTURE\_InChIKey** field (25 character abbreviated InChI for use in structure-indexing applications) has been added as a DSSTox Standard Chemical Field to all DSSTox files.

#### Log of SDF Modifications and Version/revision updates:

Date	DSSTox SDF File Names	Modifications from previous version	Additional Notes
15Oct03	EPAFHM_v1a_617_15Oct03 EPAFHM_DOP_v1a_610_15Oct03	Initial launch publication; no previous published versions.	EPAFHM is considered a "static" historical database meaning that further expansion of the database to include additional test data is unlikely. Future updates will correct reported errors provided by users or incorporate DSSTox format changes.
1Mar05	EPAFHM_v2a_617_1Mar05	New Fields: <b>INChI</b> , <b>ChemName_IUPAC</b> , <b>StudyType</b> , <b>Species</b> , <b>Endpoint</b> Corrected <b>CAS</b> for first 14 records ( <b>DSSTox_ID</b> =1-14). Corrected Structure for <b>DSSTox_ID</b> =97 (changed from penta to trichlorophenol). Addition of <b>SMILES_Parent</b> to Main file. Moved 3 <b>SubstanceType</b> ="mixture or unknown" records to end of file so that first 614 chemicals listed are those with structures.	Major format modification to include INChI, IUPAC names, and ToxML fields.  Separate "desalted" defined organic parent (DOP) file not provided. Users can easily generate DOP file by extracting "defined organic" records and converting <b>SMILES_Parent</b> to structures.
10Apr2006	EPAFHM_v3b_617_10Apr2006	Updated with new DSSTox Standard Chemical Fields and entries ( <i>revised Aug 2005</i> ). Updated InChI codes (version 1.0). Updated IUPAC chemical names (ACDLabs Name to Structure, version 9.0). Expanded "ddmmmyear" format for dates in DSSTox file names (e.g., 23Jan2006). Renamed Source-related fields to be more descriptive: <b>ChemClass FHM</b> to <b>ChemClass_FHM</b> <b>LC50</b> to <b>LC50_mg</b> <b>LC50NOTE</b> to <b>LC50_Note</b> <b>LC50RATIO</b> to <b>LC50_Ratio</b>	Numerous structure modifications and changes in stereochemical rendering throughout DSSTox data files following major quality review.  EPAFHM_v3a_617_22Oct2005: Note: earlier version of this file was provided to PubChem, with identical format to v3b but latter has undergone additional QA review and has small number of corrections/modifications.

		<p> <b>MOACONF</b> to <b>MOA_Confidence</b>  <b>MIXMOA</b> to <b>MOA_MixtureTest</b>  <b>TOXINDEX</b> to <b>ExcessToxicityIndex</b>  <b>FATS</b> to <b>FishAcuteToxSyndrome</b>  <b>BEHAVIOR</b> to <b>FishBehaviorTest</b> </p> <p>Added new field <b>LC50_mmol</b>.</p> <p>Corrected LC50 entries for ID 242 Diethanolamine from EPA Source.</p>	
15Jun2007	EPAFHM_v4a_617_15Jun2007	<p>Revised Standard Fields:</p> <p><b>DSSTox_SID</b> has been replaced by two new ID fields <b>DSSTox_RID</b> and <b>DSSTox_Generic_SID</b>.</p> <p><b>DSSTox_ID_FileName</b> has been replaced by new ID field: <b>DSSTox_FileID</b>.</p> <p>Entries in <b>TestSubstance_Description</b> field have been simplified.</p> <p>Entries in <b>ChemicalNote</b> that pertained specifically to EPAFHM have been moved to Source-Specific field: <b>Note_EPAFHM</b></p> <p><b>LC50_Ratio</b> field has been converted to a pure numeric field, with field notes moved to <b>LC50_Note</b> field.</p> <p>Entries in several Source-Specific fields have been unabbreviated or expanded:</p> <p><b>MLOGP</b>  <b>LC50_Note</b>  <b>MOA</b>  <b>MOA_MixtureTest</b>  <b>FishAcuteToxSyndrome</b>  <b>FishBehaviorTest</b></p>	
15Feb2008	EPAFHM_v4b_617_15Feb2008	<p>Total of 8 corrections:</p> <p>Added stereochem to 6 structures: RID 22047; 22072; 22126; 22163; 22233; 22301</p> <p>Corrected 1 TestSubstance_ChemName RID 22195 (lowercase n to capital N)</p> <p>Changed one substance from "representative isomer in mixture" to "representative component in mixture": RID 22306</p> <p>Modified text field entries: <b>MOA_Confidence</b></p> <p>New Standard Field added:  <b>STRUCTURE_InChIKey</b></p>	All corrections or changes to structure information noted in <b>Note_EPAFHM</b> field, searchable by version (e.g., v4b).

		Two new summary activity fields added in coordination with PubChem deposits: <b>ActivityOutcome_EPAFHM</b> <b>ActivityScore_EPAFHM</b>	

**Field and Data Counts in Older Versions of DSSTox SDF files:**

DSSTox SDF	Standard Chemical Fields	Standard Toxicity Fields	Source-specific fields	Chemical records total	STRUCTURE_ChemicalType:			STRUCTURE_TestForm_DefinedOrganic:	
					Defined Organic	Inorganic	Organo-metallic	Parent	Salt or complex
EPAFHM_v1a	13	0	12	617*					
EPAFHM_DOP_v1a	15	0	12	610					
EPAFHM_v2a	16	3	12	617*	610	1	3	589	15

\*One pair of replicate parent structures (salt or complex) and one pair of replicate 2D structures (cis and trans) exist in the database.

<b>EPAFHM SDF Content</b>	<b>Totals_v3b</b>	<b>Totals_v4a</b>	<b>Totals_v4b</b>
# Records	617	617	617
DSSTox Standard Chemical Fields	18	18	19
DSSTox Standard Toxicity Fields	3	3	3
EPAFHM Source Fields	13	14	16
Total # Fields	34	35	38
<b>Chemical Content</b>	<b>Counts_v3b</b>	<b>Counts_v4a</b>	<b>Counts_v4b</b>
<b>STRUCTURE_ChemicalType:</b>			
defined organic	613	613	613
inorganic	1	1	1
organometallic	3	3	3
no structure	0	0	0
<b>STRUCTURE_TestForm_ DefinedOrganic:</b>			
parent	591	591	591
complex	11	11	11
salt	13	13	13
salt complex	2	2	2
<b>TestSubstance_Description:</b>			
single chemical compound	610	609	609
defined mixture or formulation	7	* (NA)	* (NA)
undefined mixture	0	* (NA)	* (NA)
macromolecule	0	0	0
mixture or formulation	* (NA)	8	8

\* (NA) = field entry not applicable for DSSTox file version indicated